JOHNS HOPKINS
BLOOMBERG
SCHOOL of PUBLIC HEALTH

When Time Is of Interest: The Case for Survival Analysis

John McGready
Johns Hopkins University

Lecture Topics

- Why another set of methods?
- Event times versus censoring times
- Estimating the survival curve—the Kaplan Meier method
- Statistically comparing survival curves

JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Section A

Motivating the Need

Survival Analysis

- Statistical methods for the study of time to an event

- Accounts for . . .
  - Time that events occur
  - Different follow-up times
  - Loss to follow-up

HIV Progression among IVDUs

- From article* abstract:
  - "**Objectives**: We sought to examine whether there were differential rates of HIV incidence among aboriginal and non-aboriginal injection drug users in a Canadian setting."
  - "**Methods**: Data were derived from two prospective cohort studies of injection drug users in Vancouver, British Columbia . . . we compared HIV incidence among aboriginal and non-aboriginal participants."
  - "**Results**: Aboriginal ethnicity was independently associated with elevated HIV incidence."

Notes: * Wood, E., et al. (2003). Burden of HIV infection among aboriginal injection drug users in Vancouver, British Columbia, *American Journal of Public Health*. 98: 3.

HIV Progression among IVDUs

- From article text:
  - "Participants were eligible for our study if they were recruited between May 1996 and December 2005."
  - "As previously described, the date of HIV seroconversion was estimated by using the midpoint between the last negative and the first positive antibody test results. Participants who remained persistently HIV seronegative were censored at the time of their most recent available HIV antibody test result prior to December 2005." (end of study)

Source: Wood, E., et al. (2003). Burden of HIV infection among aboriginal injection drug users in Vancouver, British Columbia, *American Journal of Public Health*. 98: 3.

HIV Progression among IVDUs

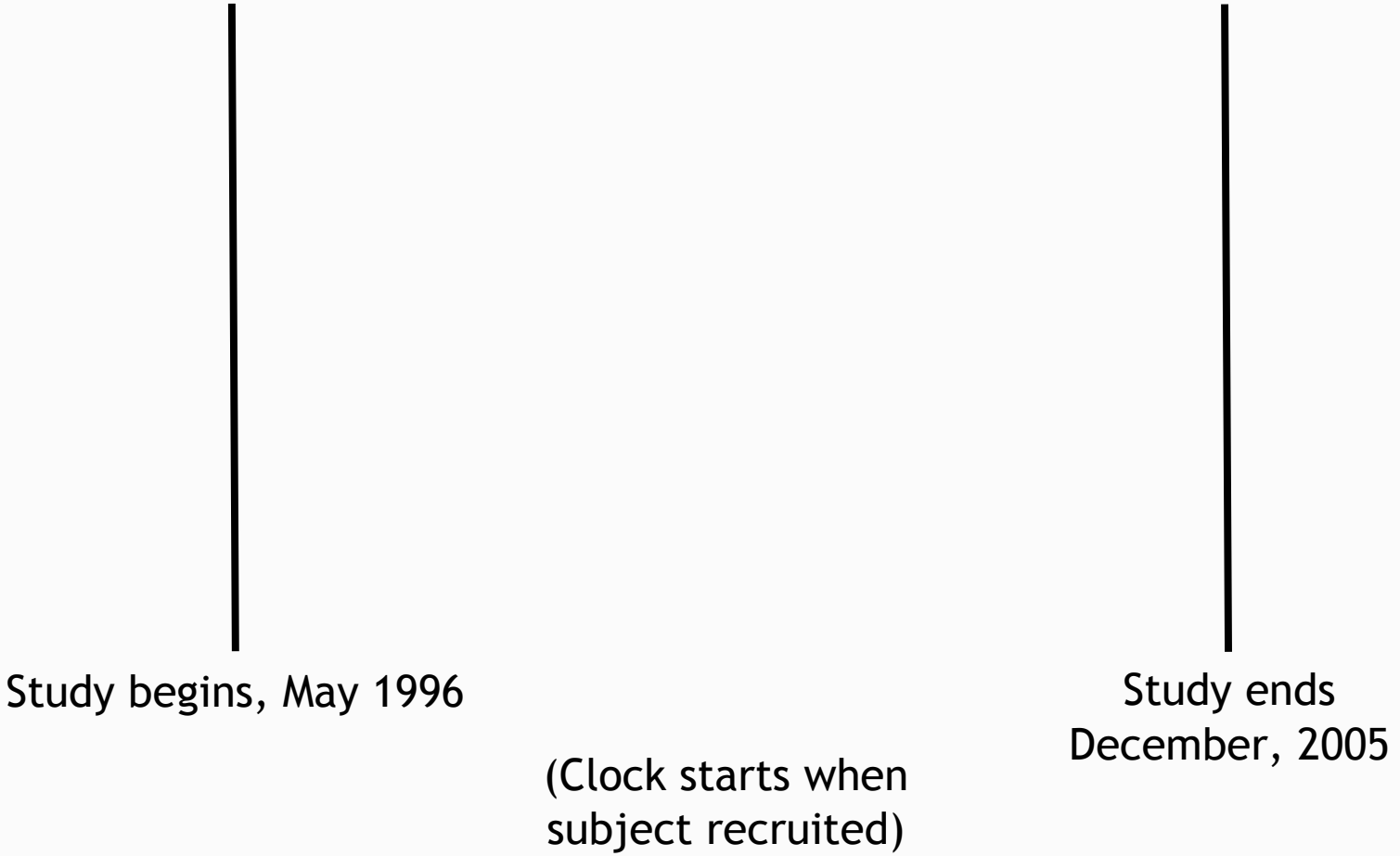
- Event of interest: HIV seroconversion

- Time frame for tracking HIV seroconversion among participants who were HIV negative at time of enrollment
  - "Clock" starts at time of enrollment"
  - "Clock" stops at either:
    - Seroconversion (event observed)
    - End of study (no event observed)
    - Loss to follow-up prior to seroconversion (no event observed)

- Researchers interested in both frequency of event AND time to event

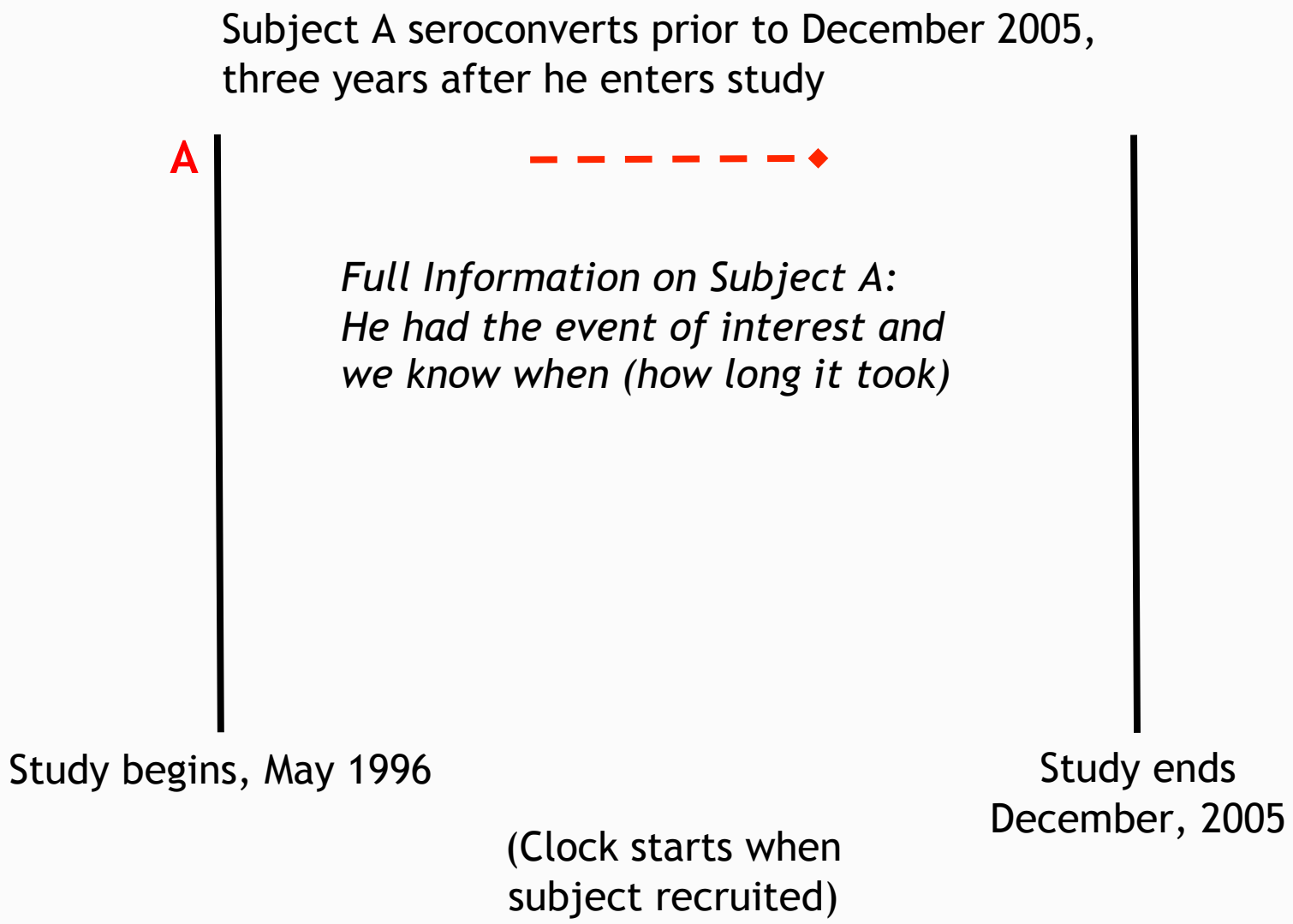
HIV Progression among IVDUs

- Graphic of possible scenarios



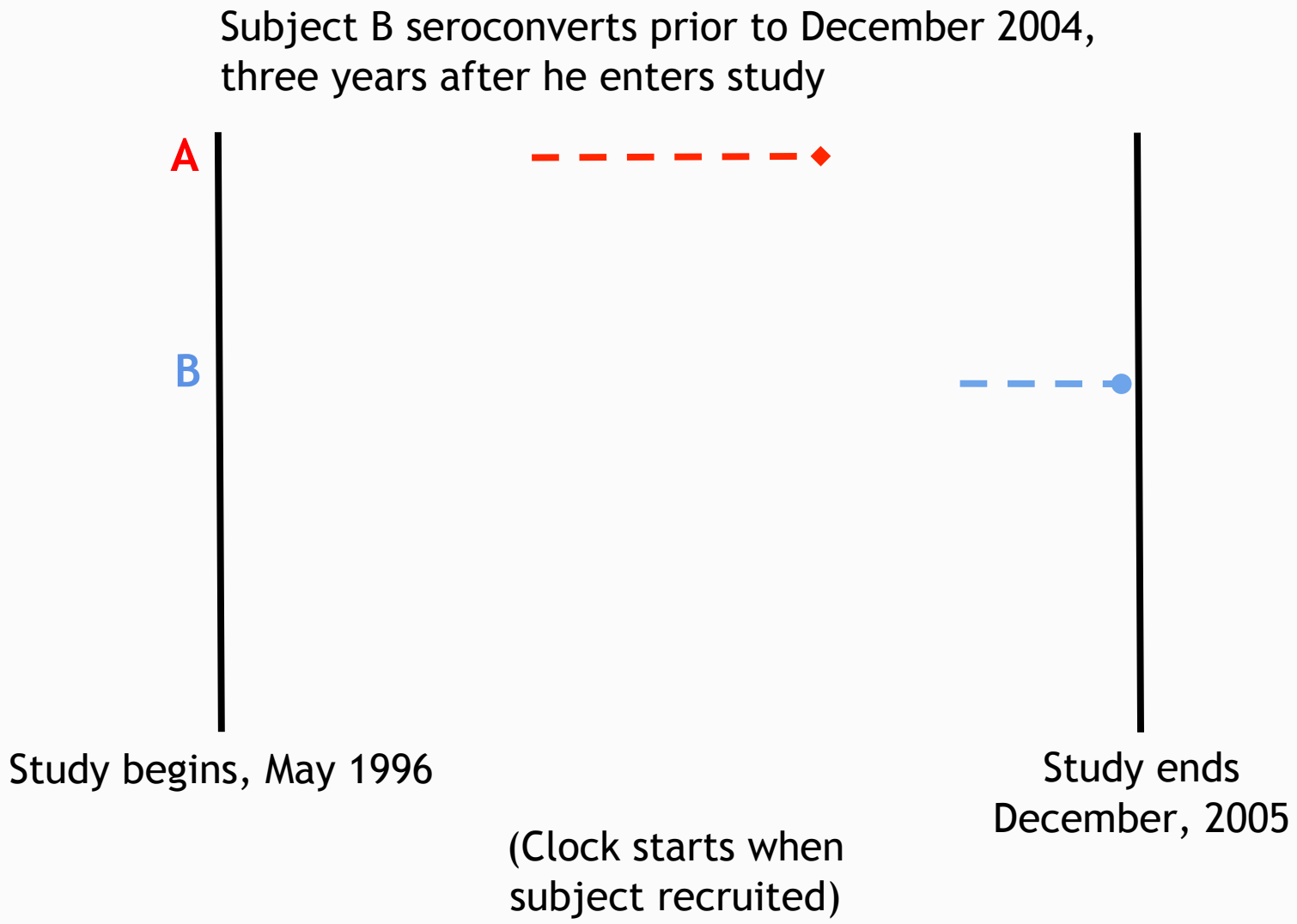
HIV Progression among IVDUs

- Graphic of possible scenarios



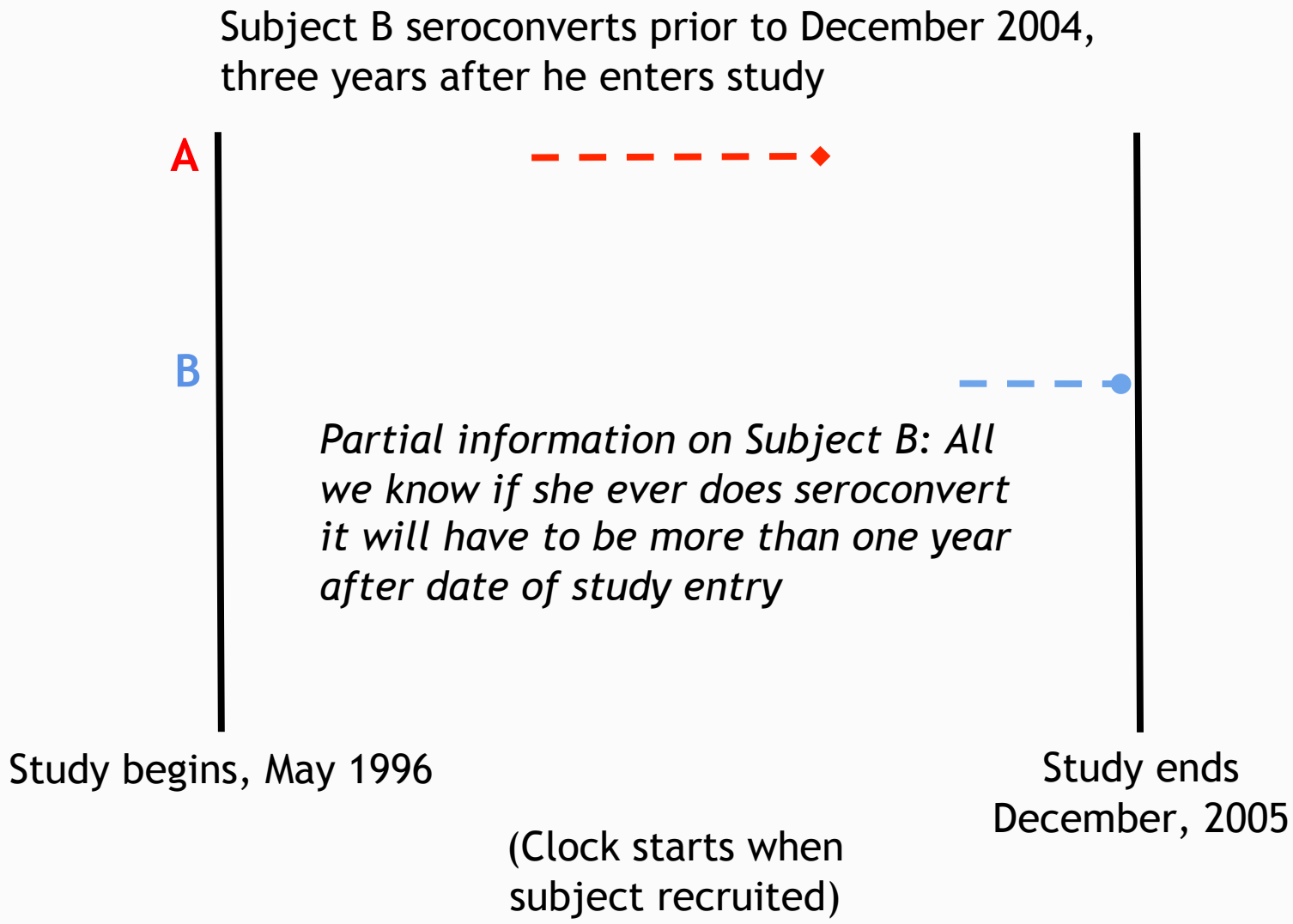
HIV Progression among IVDUs

- Graphic of possible scenarios



HIV Progression among IVDUs

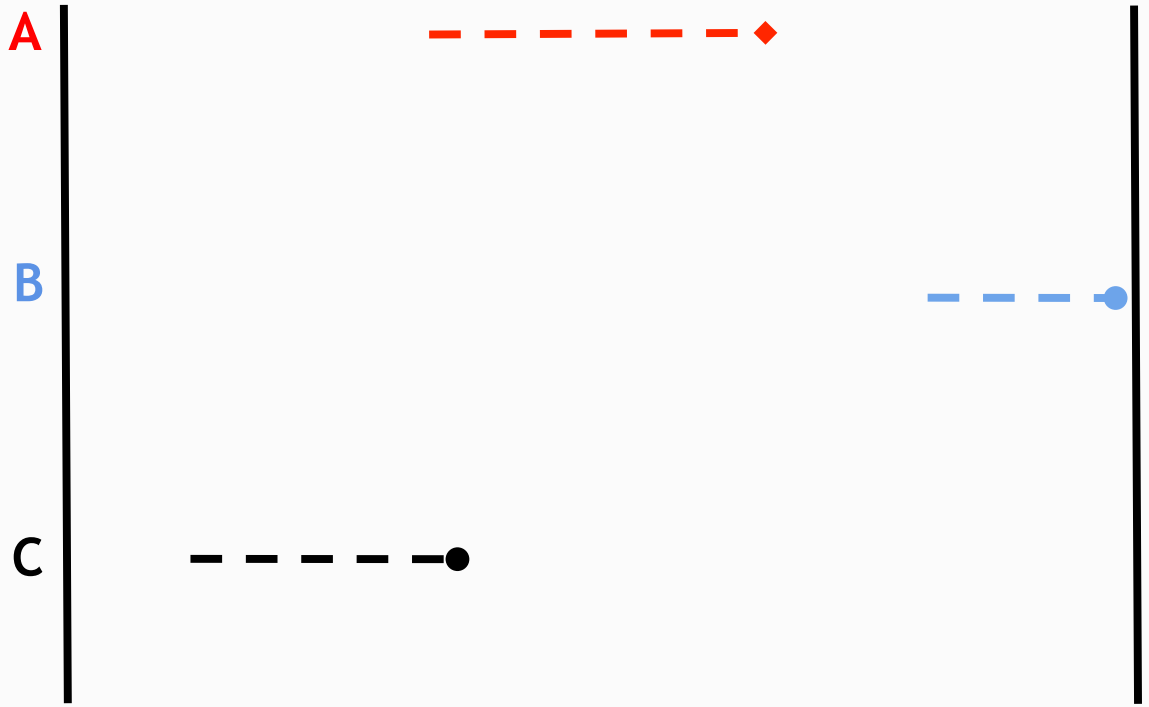
- Graphic of possible scenarios



HIV Progression among IVDUs

- Graphic of possible scenarios

Subject C enters in November 1997; He has all negative HIV tests until last follow-up visit in November 1999



Study begins, May 1996

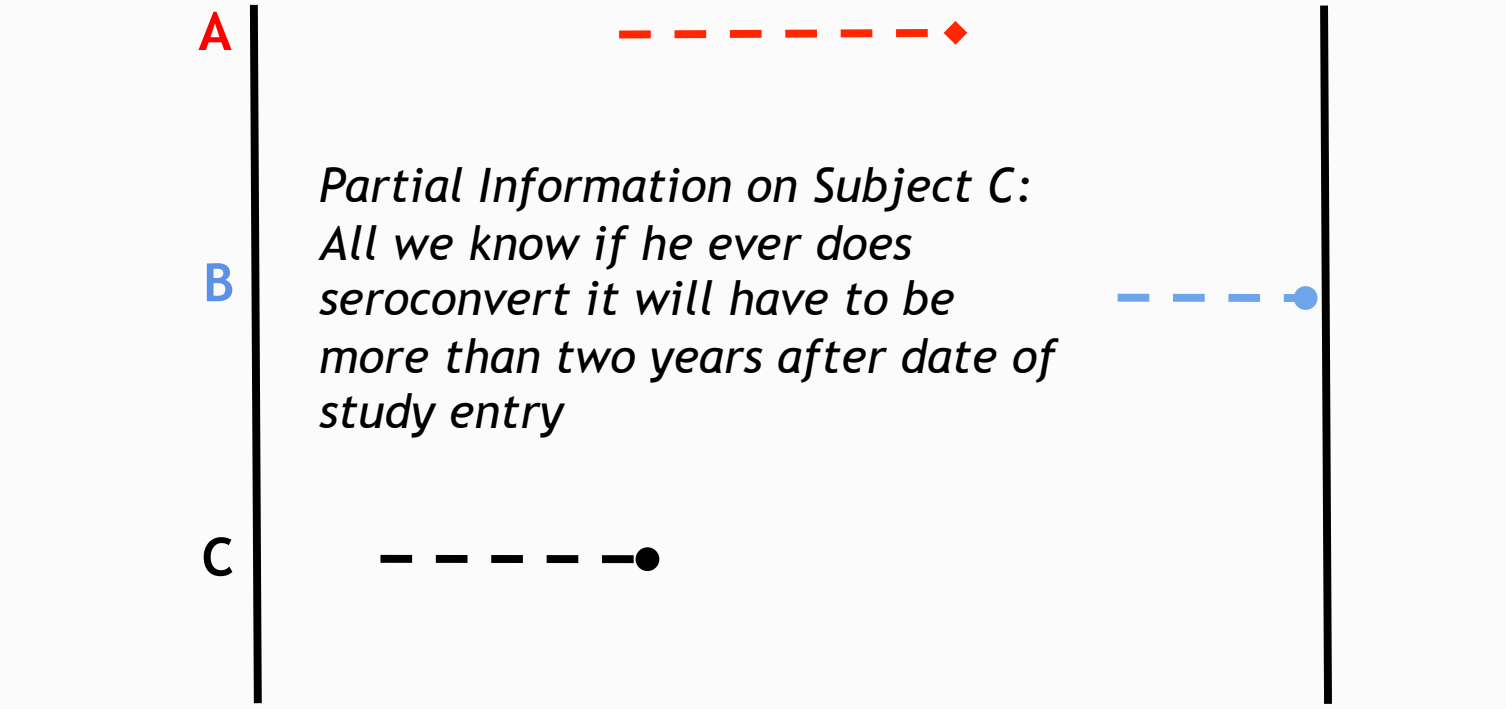
(Clock starts when subject recruited)

Study ends December, 2005

HIV Progression among IVDUs

- Graphic of possible scenarios

Subject C enters in November 1997; He has all negative HIV tests until last follow-up visit in November 1999



Study begins, May 1996

(Clock starts when subject recruited)

Study ends December, 2005

Chemotherapy Example

- Suppose we have designed a study to estimate survival after chemotherapy treatment for patients with a certain cancer
- Patients received chemotherapy between 1990 and 1994 and were followed until death or the year 2000, whichever occurred first
- In this study the event of interest is death
- The time clock starts as soon as the subject finishes his/her chemotherapy treatments

Chemotherapy Example

- Three results from study:
  - Patient one enters in 1990, dies in 1995: patient one survives five years
  - Patient two enters in 1991, drops out in 1997: patient two is lost to follow-up after six years
  - Patient three enters in 1993 and is still alive at end of study: patient three is still alive after seven years

Why Is Survival Analysis Tricky?

- Patient:
  - 1: 1990 → 1995 5 years
  - 2: 1991 → 1997 6+ years
  - 3: 1993 → 2000 7+ years

- Patients two and three are called censored observations

- We need a method which can incorporate information about censored data into an analysis

Interested in Time: Why Not Treat as Continuous

- Patient:
  - 1: 1990 → 1995 5 years
  - 2: 1991 → 1997 6+ years
  - 3: 1993 → 2000 7+ years

- Suppose we wanted to estimate the mean time to death for the three patients listed above: suppose we average the three death/ censoring times

- This average would systematically underestimate the average of the three persons, because two of the three numbers are underestimates of time to death after finishing chemotherapy

Interested in Occurrence of Event: Binary?

- Event of interest is binary
  - Why not just summarize total proportion who had the event before the end of study, treating those censored as "non-events"
  - Suppose we have designed a study to compare survival after two different chemotherapy treatments for patients with a certain cancer
  - Patients randomized to one of two chemotherapy groups: after assignments, received chemotherapy between 1990 and 1994 and were followed until death or the year 2000, whichever occurred first

Interested in Occurrence of Event: Binary?

- At end of study, 40% of patients in each of the two chemotherapy groups had died
  - Exactly the same proportion (do we even need a p-value?)
  - Does this show that neither treatment is “superior” in terms of prolonging survival?

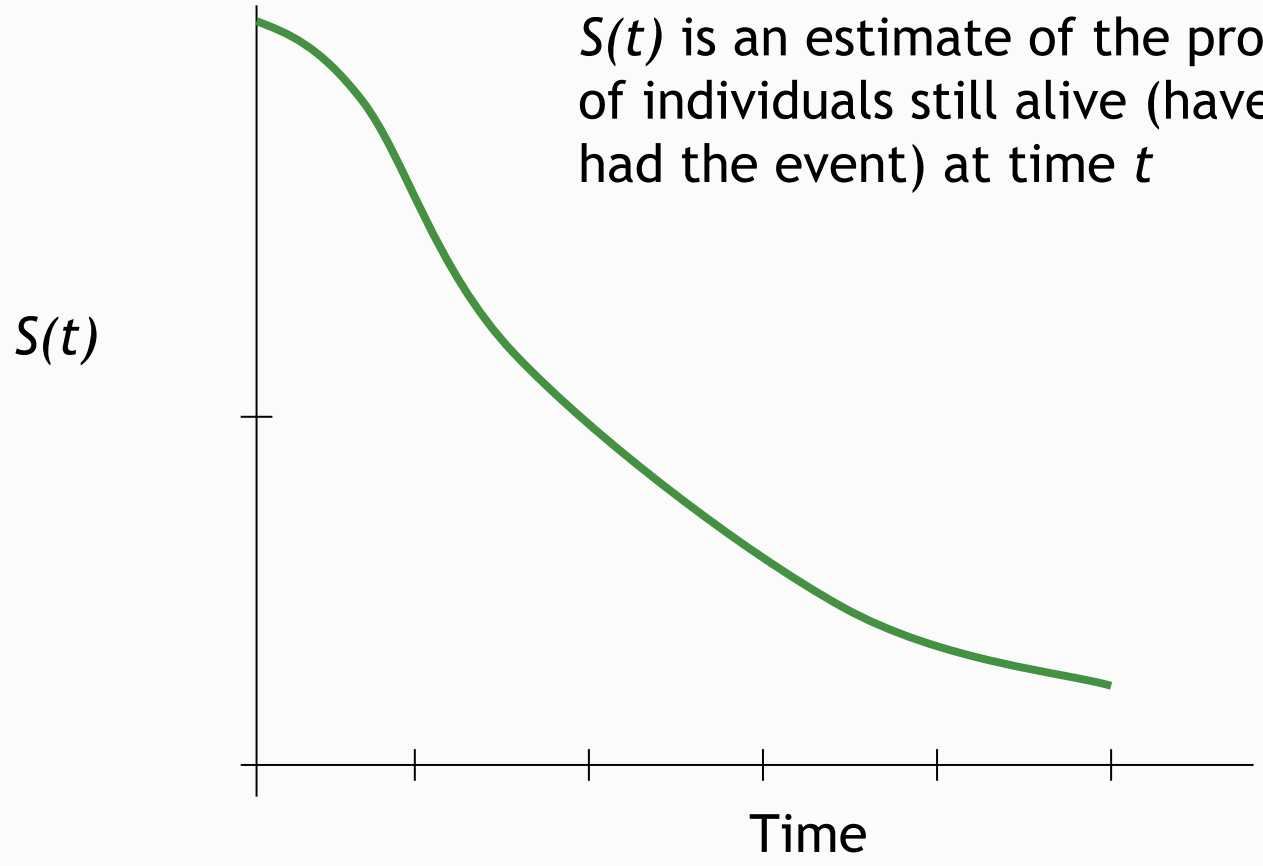
- Suppose in the first chemotherapy group, most of the 40% died within a year of stopping the treatment; in the second group, most of the 40% died between five to six years after stopping treatment:
  - Timing of the event is very different between the two groups even though the end percentages are similar

Another Method Needed

- Another method is needed to analyze time to event data in the presence of censoring
- This method needs to utilize time in its analysis, but also differentiate between event times (full time information) and censoring times (partial time information)
- This method will produce a summary statistic that captures both the binary portion (event y/n) and the time portion of the “story”

Summary Statistics

- The method we will discuss in the next section produces the following "summary statistic" for a sample of time-to-event data
  - The survival curve



S(t) is an estimate of the proportion of individuals still alive (have not had the event) at time *t*